

Plan de formation DS
Disponibilité : 30/09/24

Module 1: Introduction

- 1.1 Qu'est-ce que la science de données ?
 - 1.1.1 Définition, importance et applications
 - 1.1.2 Rôles et responsabilités d'un data scientist.
- 1.2 Outils et langages
 - 1.2.1 Présentation de python, Jupyter Notebook, R
 - 1.2.2 Installation et configuration des environnements
 - 1.2.3 Parler de Colab et Kaggle.

Module 2: Rappel des bases de python

- 2.1 Variables.
- 2.2 Lists, Dictionnaires, Ensembles, Tuples, ...
- 2.3 Fonctions, Modules, classes
- 2.4 Boucles.
- 2.5 Structures de contrôle.
- 2.6 Gestion des erreurs (opt.).
- 2.7 Exercice de consolidation.

Module 3: Numpy

- 3.1 Petite introduction.
- 3.2 Les tableaux Numpy (ndarray)
- 3.3 Manipulation des tableaux (ex: indexation, rank et shape, ..).
- 3.4 Avantages de Numpy.
- 3.5 Quizz de consolidation.
- 3.6 Exercice de consolidation.

Module 4: Manipulation des données avec pandas

- 4.1 Historique et définition
- 4.2 Où trouver les données ?
- 4.3 Chargement des données.
- 4.4 Plusieurs manières de créer un dataframe.
- 4.5 Séries vs Dataframes.
- 4.6 Series vs tableaux Numpy.
- 4.7 Filtrage, tri, agrégation, jointure.
- 4.8 Nettoyage des données (gestion des valeurs manquantes, doublons, etc...).
- 4.9 Statistiques descriptives (moyenne, médiane, écart-type, ...).
- 4.10 Quizz de consolidation.
- 4.11 Min-projet de consolidation

Module 5: Dataviz avec Matplotlib et Seaborn

- 5.1 Graphiques (Histogrammes, barplots, boxplots, scatter, pie, ...)
- 5.2 Quizz de consolidation pour valider les acquis
- 5.3 Mini-projet de consolidation sur la dataviz.

Module 6: Machine Learning (ML)

- 6.1 Introduction au ML.
- 6.2 Types d'algorithmes du ML.
- 6.3 Compromis biais vs variance (overfitting vs underfitting).
- 6.4 Outils d'implémentation des modèles et évaluation (Scikit-Learn).
- 6.5 Régression linéaire
 - 6.5.1 Intuition et choses importantes à savoir
 - 6.5.2 Implémentation par un cas pratique (modélisation et évaluation).
- 6.6 Régression logistique
 - 6.6.1 Intuition et choses importantes à savoir
 - 6.6.2 Différence avec la R. Linéaire
 - 6.6.3 Implémentation par un cas pratique (modélisation et évaluation).
- 6.7 Validation croisée (cross validation).
- 6.8 Arbres de décision
 - 6.8.1 Intuition et choses importantes à savoir
 - 6.8.2 Implémentation par un cas pratique (modélisation et évaluation).
- 6.9 Méthodes d'ensembles (ex: Random Forest)
 - 6.9.1 Intuition et choses importantes à savoir
 - 6.9.2 Différence avec un arbre de décision
 - 6.9.3 Implémentation par un cas pratique (modélisation et évaluation).
- 6.10 Envisager d'autres méthodes (ex: LazyPredict).
- 6.11 Quizz de consolidation.

Module 7: Projet Final

- 7.1 Présentation du projet et de ses objectifs.
- 7.2 Présentation des données et leur description.
- 7.3 EDA/AED
 - 7.3.1 Partie 1 : nettoyage, traitement des données
 - 7.3.2 Partie 2 : visualisation, transformation,
- 7.4 ML
 - 7.4.1 Sélection des attributs et partitionnement des données.
 - 7.4.2 Construction des modèles.
 - 7.4.3 Évaluation des modèles
 - 7.4.4 Choix du meilleur modèle et le sauvegarder.
- 7.5 Déploiement
 - 7.5.1 Intégration avec Flask - Frontend
 - 7.5.2 Intégration avec Flask - Backend

Module 8: Conclusion

- 8.1 Mots de fin et aller plus loin